Next-generation sequencing and bioinformatics in rare movement disorders

Michael Zech[1,2,3] and Juliane Winkelmann[1,2,4,5*]*

[1]Institute of Human Genetics, School of Medicine, Technical University of Munich, Munich, Germany
[2]Institute of Neurogenomics, Helmholtz Zentrum München, Munich, Germany
[3]Institute for Advanced Study, Technical University of Munich, Lichtenbergstrasse 2a, 85748 Garching, Germany
[4]Munich Cluster for Systems Neurology, SyNergy, Munich, Germany
[5]DZPG, Deutsches Zentrum für Psychische Gesundheit, Munich, Germany

*Correspondence to:
PD Dr. Michael Zech and Prof. Dr. Juliane Winkelmann
Institute of Human Genetics, School of Medicine, Technical University of Munich, Trogerstraße 32, 81675 Munich, Germany
Institute of Neurogenomics, Helmholtz Zentrum München
Deutsches Forschungszentrum für Gesundheit und Umwelt (GmbH)
Ingolstädter Landstraße 1, 85764 Neuherberg, Germany
E-mail: michael.zech@mri.tum.de, juliane.winkelmann@tum.de
Phone: 0049-89-3187-1884
Fax: 0049-89-3187-3297

Abstract

The possibility to sequence entire exomes and genomes has revolutionized the efficacy of molecular testing in rare movement disorders. Genomic sequencing is now becoming an integral part of routine diagnostic workflows for these heterogenous phenotypes. However, significant challenges are emerging in interpretation of the extensive amounts of genomic variant information being generated. In this Persepctive, we outline multidimensional strategies to genetic diagnosis in patients with rare movement disorders. We examine bioinformatic tools and computational metrics that have been developed to facilitate accurate prioritization of disease-causing variants. We highlight community-driven data-sharing and case-matchmaking

platforms, designed to foster the discovery of new gene-phenotype relationships. Finally, we consider the role of multi-omics data integration for optimization of the diagnostic success, whereby combined genomic, epigenetic, transcriptomic, and/or proteomic profiling can enable more holistic evaluation of variant effects. Together, the discussed approaches offer paths to improved understanding of the genetic basis of rare movement disorders.

Introduction

"Next-generation" DNA sequencing (NGS) assays have revolutionized the comprehensiveness of human genetic analysis by allowing simultaneous screening for variants in hundreds of disease-related genes[1]. NGS-based massive parallelization of sequencing reactions can determine the entire nucleotide sequence of an individual`s genome in a single analysis-instrument run for less than 1,000$[2]. The technique is particularly suitable for molecular studies of heterogeneous Mendelian conditions, enabling specific diagnoses to be made across a wide range of phenotypes including movement disorders[1-3].

Movement disorders form a vast category of neurological diseases with multifactorial complex genetic background on one end of the spectrum and monogenic causation on the other end, which are frequently characterized by progressive disability[4]. Many subgroups exist, defined by variable expressions of ataxia, chorea, dystonia, myoclonus, Parkinsonism, tremor, as well as phenotypically mixed syndromes, which are often individually rare and hard to categorize on clinical grounds[4]. NGS has been instrumental in identifying the monogenic causes of rare movement disorders at a broad scale[3,5-7] and in establishing an expanding catalogue of new genotype-phenotype relationships[8,9]. For example, the clinical feature

“dystonia“ may be related to monoallelic or biallelic variants in >500 genes, as currently documented in the Online Mendelian Inheritance in Man (OMIM) database[10]. With many of the associated diseases, often reported in only a few cases worldwide, the patient`s clinician may be unfamiliar. In such cases, the precise molecular diagnosis can unlock important information from the literature that may be fundamental to advising on optimal management or offering access to disorder-specific support organizations[1,2,5].

Although a substantial number of genomic datasets have been produced across different movement-disorder indications since the commercial availability of NGS in 2011, the diagnostic rates today cap at ~20-50%[11-15]. A major hinderance to a relevant increase in molecular etiological yield is our inability to interpret a significant proportion of generated sequencing information[1,2]. Individual genomes contain many thousands of ultra-rare and so-called “private“ variants, i.e., sequence changes that are found exclusively in one single studied individual. Assigning clinical relevance to such private mutations remains a difficult challenge, even when they are discovered in known disorder-associated genes[1,2]. Limitations in discovery power of NGS approaches are especially evident in the field of movement disorders because of marked contributions of variable expressivity and reduced penetrance, as well as high levels of allelic heterogeneity[8,9]; movement disorder-causing variations encompass a diverse spectrum of mutation types, including substitutions, deletions or duplications of single nucleotides, multi-nucleotide insertions and deletions, structural variants, repeat expansions, as well as mitochondrial DNA alterations[10,16,17].

To address the interpretative challenge presented by NGS, powerful sets of computational methods have been developed, supporting the identification and prioritization of disease-associated variants and genes[18-21]. However, it is increasingly difficult for movement-disorder specialists to oversee meaningful application of these

analytic algorithms. Moreover, there is widespread recognition that the NGS diagnostic process will significantly benefit from integration of other “omics“ data, such as epigenetic signatures, transcriptomics, and proteomics[22,23]. Is it time to implement multi-omics diagnostics into routine care of movement disorders? Herein, we highlight genomic analysis strategies and bioinformatic variant-prioritization approaches in the context of rare hereditary movement disorders. We describe the principles of exome- and genome-wide sequencing with a focus on variant-detection tools that are most relevant to movement disorders. Then, we discuss computational metrics and software designed to facilitate the clinically oriented filtering of variants. In addition, we outline the necessity of large-scale data sharing including online case-matchmaking initiatives and data intregration between clinical care and research to improve diagnosis. Finally, the promise of multi-omics studies is examined.

Main text

Genomic sequencing and bioinformatic workflow

The main unbiased NGS analysis techniques are whole-exome sequencing (WES) and whole-genome sequencing (WGS) (**Figure 1**)[1,2,24]. WES targets the entire protein-coding regions (~20,000 genes), comprising 1-2% of the human genome. The test is highly efficient in detecting disease-associated mutations in exonic and nearby splice-site sequences, which are currently thought to harbor the majority (~85%) of known pathogenic genomic variations[2].

The diagnostic yield of WES in movement disorders has been extensively investigated across diverse phenotype expressions and cohorts[11]. In the broad group of ataxias, for example, it has been demonstrated that ~23-52% of patients can receive a specific diagnosis by exome-wide variant profiling[15,25]. WES-driven discovery has also resulted in the identification of numerous new disease genes, as exemplified in

the description of >10 previously undefined monogenic etiologies for isolated dystonia between 2015 and 2023[26,27]. Typically, the detection rates of causative variants vary according to patient characteristics in each movement-disorder category, with generally higher chances of finding diagnoses in pediatric subjects with multisymptomatic manifestations than in less complex, often multifactorial adult cases[7,14,28]; however, WES has also been an invaluable tool in deciphering broad phenotypic spectra related to the same genetic basis, for example in *GNAO1*-linked conditions, where presentations of both infantile dyskinetic encephalopathy and late-onset focal abnormal movements have been revealed[29,30].

Despite its widespread implementation as a diagnostic tool in movement disorders, WES suffers from two key limitations[1,2]: (i) inconstant depth of sequencing coverage focused on exons hinders comprehensive detection of some clinically relevant mutation types such as certain structural variants; and (ii) sequence alterations in the ~98-99% of noncoding proportions of DNA cannot be examined. These drawbacks can be overcome by adopting a WGS approach. WGS significantly reduces the likelihood of missing disease-related variants by offering uniformity of target coverage and providing analytic access to nearly all of the ~3 billion nucleotides in a patient`s genome[31]. The benefits of WGS over WES have not yet been systematically explored in the context of rare movement disorders, but pilot studies on large heterogeneous disease populations document diagnostic uplifts and the method`s effectiveness[32,33].

Both the output from WES and WGS requires computationally sophisticated processing workflows, given that extensive amounts of data are generated (data storage requirements in the patebyte range for larger WES/WGS collections)[1,2,20]. Up to 30,000 variants can be found in an individual exome, whereas WGS usully yields ~3-4 million variant positions that differ from a reference genome[1,2]. Dedicated

bioinformatic pipelines have an essential role in the genetic laboratory, starting with WES/WGS raw-data mapping and the calling of variants. The BWA-GATK framework (Broad Institute) is considered the present gold-standard for identification of single-nucleotide variants (SNVs) and short (1-50bp) insertions and deletions (indels)[20,34], which appear to represent the largest currently recognized set of movement disorder-causing mutations. Several NGS studies of patients with ataxia, dystonia, mixed hyperkinetic syndromes and other rare movement disorders have shown that SNVs and indels accounted for the majority (~85-95%) of molecular diagnoses[6,12,14,15], although these numbers may be biased due to incomplete assessment of other variant types, especially in earlier work.

WES/WGS data can also be exploited to assess copy-number variations (CNVs), for which an increasing battery of detection algorithms is being developed[35,36]. These tools that identify CNVs in short reads from NGS machines can be coverage-based callers (e.g., ExomeDepth, CNVnator) or callers using integrated paired-end and split-read analysis strategies (e.g., DELLY, Manta)[37,38], capable of detecting deletion and duplication events with high sensitivity and specificity. CNVs represent a relevant class of genomic alterations contributing to movement-disorder manifestations, and with the growing adaption of CNV-screening tools in NGS pipelines, we are witnessing intriguing “new“ roles for “old“ deletion syndromes in movement disorders such as 22q11.2 microdeletions (DiGeorge syndrome) in Parkinsonism and hyperkinetic phenotypes[39,40]. Importantly, WGS offers the opportunity to screen genomes for a range of structural variations beyond those identifiable by WES[41]; for example, WGS has been successful in uncovering a pathogenic inversion disrupting *QDPR*[42], a gene implicated in tetrahydrobiopterin deficiency-related movement disorders. Most CNV-calling algorithms present restrictions with the discovery of small CNVs, particularly 1- or 2 exon-containing

CNVs, because these events are identified as single aberrant signals in the data with often suboptimal intensity and unwanted noise, creating challenges in routine clinical applications due to appearance of both false negatives and false positives[43].

Another important development for WES/WGS data analysis in patients with movement disorders constitutes the introduction of methods that allow scrutiny of mitochondrial DNA mutations and pathologic repeat expansions[44-46]. For the first class, so-called “off-target reads“ generated during standard WES/WGS experiments and processed in the bioinformatic pipeline can be reliably utilized for molecular diagnosis[44]. Bespoke processing workflows in genetic laboratories integrate off-capture sequencing results that derive from the enrichment of DNA fragments outside the intended target regions, including the mitochondrial genome, thereby increasing the diagnostic utility of WES/WGS[47]. A recent retrospective evaluation of 11,424 WES datasets reported detection of pathogenic mitochondrial DNA variants in 11 individuals, including patients with ataxia, dystonia, and myoclonus[48]. Despite their practical applicability, off-target read-based mitochondrial-DNA variant screens can only be performed in WES/WGS studies for determining diagnoses when the capture-target kit supports the interrogation of the mitochondrial genome (e.g., in the form of a spike-in panel with the core nuclear exome)[49]. For the second class, a specific tool named “ExpansionHunter Denovo“[50] is gaining popularity, which is suitable for performing hypothesis-free, genome-wide repeat profiling with diagnostic accuracy. In families with late-onset cerebellar ataxia, a comparatively prevalent but often genetically intractable syndrome, the tool has recently been applied to pinpoint the presence of a new repeat expansion disorder (ATX-*FGF14*)[51]. Furthermore, a systematic assessment of the test performance of ExpansionHunter-supported repeat-expansion profiling for 13 neurological disorders caused by these mutation types showed that WGS can distinguish between expanded and non-expanded alleles with 97.3%

sensitivity and 99.6% specificity[46]. On the other hand, WES has significantly restricted abilities to find causative expanded sites in patients with repeat expansion-associated movement disorders because the method cannot calculate the size of alleles larger than the commonly used read length of 100bp and because of the missing capture of non-coding parts of the disease-related genes (e.g., intron 1 of *FXN* harboring the Friedreich ataxia-linked GAA triplets)[2]. WES-based molecular diagnostics of movement disorders may thus demand additional testing for pathogenic repeats on alternative platforms, depending on the phenotypic characteristics of the examined patient.

Finally, other rare complex mutational events underlying monogenic movement disorders have begun to be unraveled in NGS data by modern analytic methodologies, including mobile element insertions in *NKX2-1*-linked childhood-onset chorea[52]. In daily practice, genetic data analysts greatly benefit from simultaneous integration of multiple independent variant callers in their pipelines and these systems should optimally offer the possibility to incorporate newly emerging tools once their diagnostic sensitivities and specificities have been validated.

Variant prioritization and pathogenicity assessment

Genomic sequencing, particularly WGS, produces an abundance of variant data, posing a challenge to find out which of the sequence changes is causally implicated in a patient`s phenotype[1,2]. To address this diagnostic bottleneck, automated workflows have been developed, supporting analysts in the assessment of the pathogenic role of variants. The process follows a series of computational mutation-filtration steps that are heavily dependent on a variety of online resources and bioinformatic tools (**Figure 1**)[18-20].

A primary filtering strategy involves the cross-referencing of patient-derived variants to catalogs of variations that are found in population controls; this helps to filter out benign alterations observed in persons who are not affected by the disease[53,54]. The Genome Aggregation Database (gnomAD), storing information from >120,000 exomes and >15,000 genomes of various geographical origins, is commonly used for this purpose[21]. In movement disorders, an important caveat needs to be considered when deploying variant exclusion with gnomAD data (**Box 1**). The dataset is not depleted for alleles associated with adult neurological conditions and reduced penetrance[1]. For example, the pathogenic p.Glu303del variant in *TOR1A* responsible for autosomal dominant generalized dystonia (penetrance ~30%)[27] is present in 30 heterozygous gnomAD carriers[21]. Moreover, we increasingly observe that variants linked to newly discovered autosomal recessive movement-disorder phenotypes, such as specific *WARS2* or *SHQ1* mutations in Parkinsonism and myoclonus[55,56], are registered as homozygous alleles in gnomAD[21], highlighting the need for careful, literature-informed evaluation. Additional web-based curated reference catalogs exist for CNVs including the Database of Genomic Variants[57] and dbVar[58], hosting structural variation data from both healthy populations and individuals with clinical phenotypes.

The inheritance of variants, ideally on the basis of recognizable familial transmission patterns, should also be included in the filtering criteria[53,54]. For simplex cases, parent-patient trio analysis has proven a highly efficient strategy to reduce the analytic burden by enabling straightforward detection of *de novo* variants via a bioinformatic “subtraction“ approach[59]. Studies show that *de novo* mutational events, including recurrent hits observed in multiple patients (e.g., p.Arg418Trp in *ADCY5*-related dyskinesia)[60], constitute a major cause of early-onset movement disorders[7,14]. Notably, new challenges in the filtering of variants on the basis of documented inheritance modes in movement disorders arise from the observation that a growing

number of genes have now been associated with both dominant and recessive phenotypes[61].

A further step is to prioritize variants according to their functional consequence and assumed deleteriousness[53,54]; this integrates the utilization of powerful computational metrics and software packages[19,20]. Currently, it is advisable to focus on protein-changing variations, consisting of two broad classes: (i) predicted loss-of-function (LoF) variants, i.e., nonsense, frameshift, and splice-site alterations; and (ii) missense variants. A priori, LoF variants may be regarded as excellent candidates for disease causation because they are expected to disrupt the gene`s reading frame. However, LoF variants are abundant in the population and each genome carries ~100 such alterations[2].

To distinguish between genes in which LoF variants are tolerated from those that are LoF intolerant, the gnomAD data-based "probability of being loss-of-function intolerant" (pLI) score has been introduced (**Figure 2a**)[21]. This metric, calculated on the basis of a comparison of observed versus expected LoF variants for each gene in gnomAD subjects, provides a statistically robust method for prioritization of LoF mutations that are likely to be clinically relevant[21] and has facilitated the establishment of many gene-phenotype associations in patients with movement disorders[14]; the pLI score has strongly supported the discovery of childhood-onset dystonia caused by haploinsufficiency of *KMT2B*[62], and it has assisted in identifying genes whose LoF variants are generally considered to be non-relevant to movement-disorder traits (e.g., LoF variants in *LRRK2* do not underlie hereditary Parkinson disease[63,64]). For missense variants, a similar measure is available, the gnomAD-derived missense-*z*-score[21]; it allows to filter WES/WGS data for genes in which missense variants are significantly underrepresented in controls. Detected missense substitutions in such genes should be carefully evaluated since the probability for pathogenicity may be

high. A typical example for a movement disorder-related gene with severe missense constraint is *ATP1A3*[21], linked to dystonia-parkinsonism and infantile dyskinetic syndromes[65]. Constraint-based approaches for missense-variant prioritization have also been developed at the levels of protein domains and individual genomic positions (i.e., at the codon level)[66-68], where such mutations can occur (**Figure 2b**); these computational methods exploit the fact that certain regions in genes and their products stand out because of their mutational invariability in the general population[66,67]. It is possible to specifically screen for missense variants that map to these mutation-intolerant sites, and are thus more likely to have a deleterious impact; in *NR4A2*-associated neurodevelopmental disorder with dystonia and parkinsonism[69-71], nearly all pathogenic missense variants are located at invariant amino acid residues in a protein motif that is heavily depleted for functional variation in population controls[72]. A cautionary note is the increasing appreciation of limitations in the use of mutational constraint parameters for the interpretation of LoF and missense variant pathogenicity in rare movement disorders, as outlined in **Box 1**. Additional information for filtering missense variants includes predicted effect on protein structure and evolutionary conservation, assessable by several commonly available *in-silico* classifiers[19,20,73]. Some more recently introduced tools function as “metapredictors“ with higher positive predictive values, combining multiple outputs from different published algorithms for estimation of the functional deleteriousness of a given amino acid change[74]; these aggregate packages assist with missense-variant evaluation processes, but their results need to be interpreted in conjunction with the peculiarities of the involved disease-associated proteins (**Box 1**).

An additional stage, useful when variants were pre-filtered as described above, involves the implementation of candidate gene lists obtained on the basis of the patient`s phenotypic characteristics[53,54]. This “virtual panel“ approach narrows the list

of selected variants to those affecting genes for which an association with the presenting clinical features has been established. Recently, a regularly curated virtual gene panel catalog has been launched via a publicly available platform, PanelApp[75]. Alternatively, gene lists can be downloaded from OMIM[10], although not all entries may be up-to-date.

Finally, for any prioritized variant, a standardized framework for clinical interpretation has to be applied, as provided by the American College of Medical Genetics and Genomics (ACMG; 5-tier classification system)[76]. To reduce inter-rater variability and the risks of subjective evaluation of variants, software tools are now becoming available that provide automated ACMG guideline-based interpretative outputs for WES/WGS-filtered variants such as wInterVar[77].

Despite all the advancements in bioinformatics-driven variant categorization, a massive number of sequence changes remain of undetermined clinical significance (so-called “variants of uncertain significance“, VUS). A major controversy is whether these variants that cannot reach a consensus on disease causality should always be reported back to referring clinicians and the affected families[78]. VUS classification criteria are conservative and designed to prevent potential harm that may result from erroneous pathogenicity assignments on the basis of insufficient evidence[76]. VUS reporting can trigger manifold patients` responses and should follow careful guidelines in order to avoid enhanced medical uncertainty and negative psychosocial impact[79]. One important strategy for dealing with VUS is to implement internal reanalyses of these findings and the otherwise unresolved WES/WGS data at periodic intervals[80]. It was shown that successful reanalysis approaches of existing NGS data, including integration of updated database annotations, consideration of more detailed phenotype information, and searches for the latest published gene-disease and variant-disease relationships, can increase the diagnostic yield by ~6-47% via various measures

including VUS upgrading[81]. Reclassifications of VUS through reanalysis may be more difficult to achieve for patients from understudied geographic areas who display distinct allelic architectures with specific rare variants including founder mutations that are not registered in available reference databases; global efforts are needed to generate ancestry-specific allele datasets, as recently realized for a larger population from the Middle East[82]. As WGS becomes more widely deployed in the field of monogenic movement disorders, we will likely see an exponentially growing set of hard-to-interpret variants in non-coding genomic regions, which will further increase the quantity and complexity of VUS information. Unraveling variants that are unequivocally causal for movement disorders can also remain a difficult challenge because of the genetic heterogeneity and clinical variability associated with these conditions; in an example of extreme phenotypic pleiotropy, independent groups have recently implicated genes of the nucleotide-excision DNA repair pathway that had originally been linked to hereditary skin disorders with photosensitivity and cancer in rare movement disorders with chorea, dystonia, and ataxia[83,84].

Data sharing

Over the past years, it has become clear that genomic data produced by individual laboratories are frequently not sufficient to generate compelling evidence for causality of VUS in known disease genes or candidate variants in potential new disorder-relevant loci[85]. In rare movement disorders, only a very few individuals seen at a single institution are affected by a specific syndrome, and the vast majority of sequenced individuals in each local database do not share the same variant hits. The establishment of mutation recurrence in independent similarly affected subjects and the identification of multiple patients with variants in the same gene are required to firmly define genotype-phenotype correlations[86].

The challenge posed by VUS and the situation of having identified one single family with a promising but unconfirmed gene candidate ("N-of-1 problem") can be overcome by sharing observations regarding rare molecular findings with the broader genetics community[85,87]. Collaborative efforts in global projects have addressed this need by developing data-sharing solutions that provide centralized repositories for clinically evaluated variants as well as platforms for genotype-phenotype matchmaking[87,88]. Databases that actively catalog sequences changes according to their previously reported disease relevance include ClinVar[17] and The Human Gene Mutation Database[89]. These sources summarize information on variant pathogenicity, mostly for SNVs and short indels, that would otherwise be dispersed across the literature and private mutation compendia, allowing analysts to quickly judge the significance of WES/WGS-identified variants.

Similar clinically-centered annotation platforms are available for CNVs (DECIPHER)[90] and mitochondrial DNA variants (Mitomap)[91]. However, since these databases are human curated and sometimes filled with spurious phenotype-gene associations, misclassifications or conflicting interpretations of variants are not uncommon[1]. This situation is especially problematic in the molecular analysis of rare movement disorders because the capture of reliably interpreted variant calls is generally under-represented in these indications (as compared to, e.g., intellectual disability), which can increase the burden of diagnostic uncertainty. Therefore, it is crucial to motivate genetic laboratories focusing on movement disorders to systematically submit their sequencing results to ClinVar and other public knowledge repositories. In relation to this, alternative community-based curation platforms have been launched such as the Movement Disorder Society Genetic mutation database (MDSGene) which promote meaningful exploration of the evidence of variant pathogenicity in the context of ataxia, chorea, dystonia, and other movement-disorder

presentations[92]. Still, a lack of diversity in genomic testing among ethnic groups with significant underrepresentation of certain populations (e.g., minority groups, African ancestry populations) remains a major hurdle for understanding of shareable pathogenic variation and cross-laboratory mutation (re-)assessments in invididuals of all geographical backgrounds[93].

To increase the analytic power in rare-disease diagnostics, important genotype- and phenotype-driven matching algorithms have been set up, including the international Matchmaker Exchange (MME) service[94]. Initiated in 2013, MME introduces genetic data-sharing mechanisms and tools for phenotypic analysis that are incorporated into a federated system, acting towards the common goal of catalyzing connections between clinicians and researchers with interest in the same genes and disorders[94]. The MME network is joined by a series of connected nodes, among which GeneMatcher has emerged as one of the most widely applied tools[95]. GeneMatcher, representing data from thousands of disease subjects, is accessible to medical professionals from around the globe and has profoundly facilitated the identification of patients with similar genotypic and phenotypic profiles (**Figure 3**)[95]. With this platform, a substantive number of rare and ultra-rare movement-disorder cases have been matched, leading to characterization of many previously unrecognized disease entities[96]. For example, WES recently revealed a “private“ missense variant in *ATP5MC3*, encoding an essential component of the mitochondrial respiratory chain complex V, in an US-American family affected by dominantly inherited dystonia and spasticity[97]; because the variant had never been described before in independent cases, it qualified as a VUS and the family remained undiagnosed. The finding was entered into GeneMatcher, which ultimately yielded a “match“ through identification of the exact same mutation in a German dystonia pedigree; this resulted in the discovery of a novel mitochondrial defect-related monogenic movement disorder[97].

Several additional data-sharing initiatives that allow comparison of sequencing findings can be helpful in genetic studies of rare movement disorders such as platforms that register systematic information on *de novo* variants identified from trio-WES/WGS analyses[98]; mostly, these listed *de novo* mutations are derived from patients with neurodevelopmental diseases[98], but their consideration can be useful in movement-disorder diagnostics given that movement disorders and neurodevelopmental diseases often share a common genetic basis[14]. The establishment of interoperable national and continent-wide data hubs, offering further improved paths to the sharing of ethnically diverse genetic information in common databases, is also underway[85]. For example, the European Genome-Phenome Archive (EGA), and its German hub GHGA (German Human Genome-Phenome Archive) support deposition of genomic sequences and phenotypes, including movements disorders, to optimize data reuse and accelerate disease gene discovery[99]. Similarly, increasing efforts are now geared towards sharing pan-European rare-disease data in a systematic manner within the Solve-RD research consortium[100]. Finally, it is interesting to note that when data sharing through health professionals is unable to provide diagnostic clarity, some patient families take over responsibility and advertise their genetic information via social media to make themselves more “discoverable“[87].

Integration of multi-omics studies

DNA-level approaches such as WES and WGS are limited in their ability to clarify the significance of a large proportion of variants in disease manifestation[1,2]. Although reanalyses, predictive algorithms, and data sharing have improved prioritization and interpretation of genetic findings, the pathogenicity of VUS and alterations situated in non-coding areas can often not be confirmed or invalidated by these methodologies. Parallel assessment of additional “omics“ layers offers an opportunity to overcome

these hurdles (**Figure 4**)[22]. There is a growing body of literature linking genomic sequencing results with epigenetic, transcriptomic, and/or proteomic data to reveal pathophysiological mechanisms and uncover diagnoses in previously unresolved monogenic phenotypes[23].

At present, strategies for multi-omics data integration have little systematic application to rare movement disorders, but first studies demonstrate that they are very promising for improving diagnostic performances[101] and international collaborations have been put in place to scale their use in molecular characterization of patients with dystonia and other indications (e.g., https://www.ejprarediseases.org/). Genome-wide analysis of DNA methylation marks is able to identify biologically meaningful signals that can support the evaluation of variant effects[102]. Initially introduced in the field of neurodevelopmental diseases, these genomic “episignatures“ were shown to be especially effective in providing clues to the underlying genetic basis for conditions that are linked to genes with suspected impact on DNA-methylation status[102]. Some studies have now been undertaken to develop accurate episignature-based classifiers for variants in the dystonia-linked gene *KMT2B*, which encodes a histone methyltransferase involved in epigenetic modifications[103,104]. In one publication, a blood-derived, disorder-specific episignature on 113 DNA methylation sites was used to re-classify four VUS in *KMT2B* (three of which newly qualified as disease-causing), leading to optimizations in diagnostic outcome and therapy-relevant results given that *KMT2B*-related dystonia is highly responsive to deep brain stimulation[103]. Moreover, DNA methylation profiling appears to allow predictions of age-of-onset and disease severity in patients with *KMT2B* mutations[103].

The study of transcriptomes by RNA sequencing (RNA-seq) represents another complementary assay to WES/WGS analyses[105]. The technique examines RNA levels in an unbiased manner, both qualitatively (integrity of transcripts) and quantitatively

(amounts of expression), and can provide a broad view of transcription-related pathological events[105]. RNA-seq data can be particularly important for interpretation of non-coding variations, but also for assessment of the effects of synonymous variants that can impact splicing[22]. Most RNA-seq pilot studies performed on patients with heterogeneous rare-disease presentations aimed at deciphering the roles of uncertain WES/WGS findings by focusing on detection of splicing mutation-induced aberrant transcripts and/or aberrant expression states[105,106]. The value of exploring these types of pathologies for identification of additional diagnoses in individuals affected by movement-disorder features is beginning to emerge in a recent large-scale combined WES/WGS-RNA-seq study[107], describing missplicing and pathologically decreased expression of *TIMMDC1* as a result of deep intronic variants in patients with ataxia and dyskinetic movements.

Further diagnostically useful information in multi-omics studies can be added at the level of proteomic investigations[22]. Proteomics assists in rare-disease variant interpretation by identifying instances where VUS have resulted in abnormally down-/or up-regulated protein expression. Although still in its infancy in rare movement-disorder diagnostics, quantitative proteomics has recently been applied successfully in a child with dyskinetic epileptic encephalopathy for functional validation of candidate variants in *ATP5PO*, establishing the diagnosis and characterizing a new recessive neurogenetic syndrome[101].

Tissue-specific expression is an important aspect that needs to be considered during transcriptomic and proteomic studies[23]. While brain samples are mostly unaccessible for these purposes, a routine practice is to perform extraction from patient-derived skin fibroblasts, in which thousands of RNAs and proteins can be reliably assessed[107]. An alternative approach is the investigation of blood transcriptomes, which allows non-invasive diagnostic identification of RNA defect-

related molecular drivers of disease[108]. Optimally, the results of the different “omics“ analyses should not be evaluated separately, but processed through a unifying bioinformatic framework (“multi-omics pipeline“) that allows superimposition of all layers of information to maximize the power for functional annotation of variants. Further “omics“ methods beyond those described above, e.g., metabolomics, may also find their way into the diagnostic workflows for rare movement disorders.

Conclusions and future opportunities

The advent of NGS with its associated computational analytic tools has opened a new era in the diagnostics of rare movement disorders. Undoubtedly, WGS will become the cornerstone for molecular analysis of most, if not all patients affected by these conditions. In this scenario of broad application across diverse disease subgroups, it will be of utmost importance to establish broadly accepted standards for incorporation in daily routine and interdependent training of movement-disorder specialists and neurogenetics experts, who should work in concert to appear at the forefront of the implementation process of clinically meaningful genomic medicine that is beneficial for patients with rare movement phenotypes (**Box 2**).

Evidence-based diagnostic pathways towards a “genomic analysis-first“ approach need to be developed for individual movement-disorder indications; for example, in the field of dystonia, a predictive clinical scoring system has been proposed to incorporate genomics as an integral part of routine care[14]. Moreover, much remains to be improved in the analysis of technically difficult-to-identify mutations, VUS assessment, and the integration of information from multiple “omics“ sources in order to extract the full potential of large genomic datasets[87,88]. Concerted efforts of the neurogenomics community - composed of clinicians, human geneticists, scientists, and

bioinformaticians - will be necessary to design new or updated technologies and software that can increase the diagnostic power.

The introduction of “third-generation“ long-read sequencing offers a promising route to advanced investigation of complex genomic variations, including balanced structural variants and chromosomal rearrangements, and the method may change the way we are using WGS over time[87]. Long-read approaches, as provided by single-molecule real-time sequencing (Pacific Biosciences) and nanopore sequencing (Oxford Nanopore Technologies), can analyze genomes at the individual nucleotide level without conventional amplification steps and are thought to excel at mapping certainty and detection of mutations in repetitive DNA segments and pseudogenes[109]. The method is also effective in sequencing through large expanded repeats, providing a prospective tool for the study of repeat expansion-related movement disorders with superior performance in terms of accuracy and speed compared to traditional PCR-based strategies[110]. Sophisticated software algorithms are under investigation that attempt to support the clinical annotation of variants falling into non-coding DNA regions such as models for the prediction of splice-disrupting intronic variants and mutations altering regulatory functionality[111], representing a necessary step for improved automated interpretation of WGS data. It is also encouraging to see that genomic techniques are leading to progress in the systematic evaluation of movement disorder-associated variants with less severe phenotypic impact and reduced penetrance mechanisms including modifiers; examples include the description of *TBP* repeat expansions coexisting with pathogenic SNVs in *STUB1* in patients with ataxia, suggestive of digenic inheritance[112], and the establishment of a comprehensive database for *GBA* risk variants contributing to Parkinson disease with varying effect sizes[113]. Another important development is the introduction of scalable approaches for functional mutation-outcome measurements with translational potential for treatment:

in the field of rare monogenic *LRRK2*-associated Parkinsonism, high-throughput experimental assays have been set up to determine the biochemical consequence of any SNV identified from a patient`s genomic sequencing dataset, including VUS, paving the way for more efficient therapy trials[114]. In parallel to these advances, large-scale organization-level collaborative research initiatives are already being active to address the challenges of producing complete catalogs of monogenic phenotypes and boost mechanistic understanding of how particular mutations relate to disease biology; in the US, the NIH Undiagnosed Diseases Network aims to improve approaches to discovering the underlying etiology of undiagnosed rare conditions, implementing pipelines for genomics, multi-omics, and functional model studies[115]. The latter represent an additional essential component for the characterization of unique variants and novel gene discoveries, given that modeling in flies, worms, zebrafish, mice or patient-derived neuronal cells and organoids can yield unparalleled insights into the pathophysiological consequences of individual genotypic abnormalities[54]. A further complementary approach for evaluating rare gene defects in movement disorders in the future may be the use of systems biology, which can unveil mechanistically relevant signatures and molecular pathogenic drivers via network analyses and other computational methodology-based frameworks[116].

In these regards, we should continue to invest in the development of artificial intelligence including generative AI as well as corresponding standards for application in movement-disorder diagnostics, which will be essential for optimized prioritization of different variant types, accurate pathogenicity predictions, and widespread applicability of multi-omics analyses[88]. Correlation of these data with output from “real-world”-learning digital tools such as wearable sensors may offer additional transformative opportunities to objectively evaluate the role of certain patient-specific molecular alterations in rare movement disorders. Ongoing data-sharing activities will constitute

another driving force behind the scaling of clinically sound variant interpretations and additional disease-gene discoveries[85], while investigators should promote ethnic diversity within genomic approaches[117]. A worldwide data-sharing platform for genetic ataxias is presently being launched[118] and could serve as a blueprint for similar initiatives targeting other rare movement-disorder subtypes. Such efforts should focus on generalizability of knowledge for patients with heterogeneous demographic characteristics such as geographical origin, sex, and age. In the context of data sharing, it will also be key to put forth strategies to enhance the transfer of clinical information in the research setting, thus facilitating bidirectional integration of insights between the clinic and the scientific arena.

Ultimately, further insights into the molecular causes of rare movement disorders will yield unique opportunities for etiology-directed therapeutic interventions and “N-of-1 trials”, either by uncovering novel treatment targets or by highlighting possibilities for drug re-purposing[1]. Some inspiring examples are emerging, such as the recent demonstration of caffeine administration as a rational effective approach to the therapy of *ADCY5*-dyskinesia[119]; NGS-identified *ADCY5* mutations were shown to induce gain of protein function, a pathology that could be specifically reversed by adenosine-A2A-receptor antagonists such as the natural compound caffeine[119]. With continued progress in NGS and bioinformatics applications in rare movement disorders, we can look forward to a future where many patients could expect a precise genetic diagnosis and, hopefully, an increasing availability of personalized therapeutic agents.

References

1 Rexach, J., Lee, H., Martinez-Agosto, J. A., Nemeth, A. H. & Fogel, B. L. Clinical application of next-generation sequencing to the practice of neurology. *Lancet Neurol* **18**, 492-503, doi:10.1016/S1474-4422(19)30033-X (2019).

2 Foo, J. N., Liu, J. J. & Tan, E. K. Whole-genome and whole-exome sequencing in neurological diseases. *Nat Rev Neurol* **8**, 508-517, doi:10.1038/nrneurol.2012.148 (2012).

3 Olgiati, S., Quadri, M. & Bonifati, V. Genetics of movement disorders in the next-generation sequencing era. *Mov Disord* **31**, 458-470, doi:10.1002/mds.26521 (2016).

4 Abdo, W. F., van de Warrenburg, B. P., Burn, D. J., Quinn, N. P. & Bloem, B. R. The clinical approach to movement disorders. *Nat Rev Neurol* **6**, 29-37, doi:10.1038/nrneurol.2009.196 (2010).

5 Cordeiro, D. *et al.* Genetic landscape of pediatric movement disorders and management implications. *Neurol Genet* **4**, e265, doi:10.1212/NXG.0000000000000265 (2018).

6 Kim, M. J., Yum, M. S., Seo, G. H., Ko, T. S. & Lee, B. H. Phenotypic and Genetic Complexity in Pediatric Movement Disorders. *Front Genet* **13**, 829558, doi:10.3389/fgene.2022.829558 (2022).

7 Perez-Duenas, B. *et al.* The Genetic Landscape of Complex Childhood-Onset Hyperkinetic Movement Disorders. *Mov Disord* **37**, 2197-2209, doi:10.1002/mds.29182 (2022).

8 Marras, C. *et al.* Nomenclature of genetic movement disorders: Recommendations of the international Parkinson and movement disorder society task force. *Mov Disord* **31**, 436-457, doi:10.1002/mds.26527 (2016).

9 Lange, L. M. *et al.* Nomenclature of Genetic Movement Disorders: Recommendations of the International Parkinson and Movement Disorder

Society Task Force - An Update. *Mov Disord* **37**, 905-935, doi:10.1002/mds.28982 (2022).

10 Hamosh, A., Scott, A. F., Amberger, J. S., Bocchini, C. A. & McKusick, V. A. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res* **33**, D514-517, doi:10.1093/nar/gki033 (2005).

11 Gorcenco, S. *et al.* New generation genetic testing entering the clinic. *Parkinsonism Relat Disord* **73**, 72-84, doi:10.1016/j.parkreldis.2020.02.015 (2020).

12 Kwong, A. K. *et al.* Exome sequencing in paediatric patients with movement disorders. *Orphanet J Rare Dis* **16**, 32, doi:10.1186/s13023-021-01688-6 (2021).

13 Trinh, J. *et al.* Utility and implications of exome sequencing in early-onset Parkinson's disease. *Mov Disord* **34**, 133-137, doi:10.1002/mds.27559 (2019).

14 Zech, M. *et al.* Monogenic variants in dystonia: an exome-wide sequencing study. *Lancet Neurol* **19**, 908-918, doi:10.1016/S1474-4422(20)30312-4 (2020).

15 Sun, M. *et al.* Targeted exome analysis identifies the genetic basis of disease in over 50% of patients with a wide range of ataxia-related phenotypes. *Genet Med* **21**, 195-206, doi:10.1038/s41436-018-0007-7 (2019).

16 Martinez-Rubio, D. *et al.* Mutations, Genes, and Phenotypes Related to Movement Disorders and Ataxias. *Int J Mol Sci* **23**, doi:10.3390/ijms231911847 (2022).

17 Landrum, M. J. *et al.* ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res* **44**, D862-868, doi:10.1093/nar/gkv1222 (2016).

18 Boone, P. M., Wiszniewski, W. & Lupski, J. R. Genomic medicine and neurological disease. *Hum Genet* **130**, 103-121, doi:10.1007/s00439-011-1001-1 (2011).

19 Cooper, G. M. & Shendure, J. Needles in stacks of needles: finding disease-causal variants in a wealth of genomic data. *Nat Rev Genet* **12**, 628-640, doi:10.1038/nrg3046 (2011).

20 Pereira, R., Oliveira, J. & Sousa, M. Bioinformatics and Computational Tools for Next-Generation Sequencing Analysis in Clinical Genetics. *J Clin Med* **9**, doi:10.3390/jcm9010132 (2020).

21 Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434-443, doi:10.1038/s41586-020-2308-7 (2020).

22 Crowther, L. M., Poms, M. & Plecko, B. Multiomics tools for the diagnosis and treatment of rare neurological disease. *J Inherit Metab Dis* **41**, 425-434, doi:10.1007/s10545-018-0154-7 (2018).

23 van Karnebeek, C. D. M. *et al.* The role of the clinician in the multi-omics era: are you ready? *J Inherit Metab Dis* **41**, 571-582, doi:10.1007/s10545-017-0128-1 (2018).

24 Keogh, M. J. & Chinnery, P. F. Next generation sequencing for neurological diseases: new hope or new hype? *Clin Neurol Neurosurg* **115**, 948-953, doi:10.1016/j.clineuro.2012.09.030 (2013).

25 Coutelier, M. *et al.* Efficacy of Exome-Targeted Capture Sequencing to Detect Mutations in Known Cerebellar Ataxia Genes. *JAMA Neurol* **75**, 591-599, doi:10.1001/jamaneurol.2017.5121 (2018).

26 Keller Sarmiento, I. J. & Mencacci, N. E. Genetic Dystonias: Update on Classification and New Genetic Discoveries. *Curr Neurol Neurosci Rep* **21**, 8, doi:10.1007/s11910-021-01095-1 (2021).

27 Lange, L. M. *et al.* Genotype-Phenotype Relations for Isolated Dystonia Genes: MDSGene Systematic Review. *Mov Disord* **36**, 1086-1103, doi:10.1002/mds.28485 (2021).

28 Posey, J. E. *et al.* Molecular diagnostic experience of whole-exome sequencing in adult patients. *Genet Med* **18**, 678-685, doi:10.1038/gim.2015.142 (2016).

29 Feng, H. *et al.* Movement disorder in GNAO1 encephalopathy associated with gain-of-function mutations. *Neurology* **89**, 762-770, doi:10.1212/WNL.0000000000004262 (2017).

30 Wirth, T. *et al.* Highlighting the Dystonic Phenotype Related to GNAO1. *Mov Disord* **37**, 1547-1554, doi:10.1002/mds.29074 (2022).

31 Turro, E. *et al.* Whole-genome sequencing of patients with rare diseases in a national health system. *Nature* **583**, 96-102, doi:10.1038/s41586-020-2434-2 (2020).

32 Investigators, G. P. P. *et al.* 100,000 Genomes Pilot on Rare-Disease Diagnosis in Health Care - Preliminary Report. *N Engl J Med* **385**, 1868-1880, doi:10.1056/NEJMoa2035790 (2021).

33 Bertoli-Avella, A. M. *et al.* Successful application of genome sequencing in a diagnostic setting: 1007 index cases from a clinically heterogeneous cohort. *Eur J Hum Genet* **29**, 141-153, doi:10.1038/s41431-020-00713-9 (2021).

34 Di Resta, C., Pipitone, G. B., Carrera, P. & Ferrari, M. Current scenario of the genetic testing for rare neurological disorders exploiting next generation sequencing. *Neural Regen Res* **16**, 475-481, doi:10.4103/1673-5374.293135 (2021).

35 Pfundt, R. *et al.* Detection of clinically relevant copy-number variants by exome sequencing in a large cohort of genetic disorders. *Genet Med* **19**, 667-675, doi:10.1038/gim.2016.163 (2017).

36 Royer-Bertrand, B. *et al.* CNV Detection from Exome Sequencing Data in Routine Diagnostics of Rare Genetic Disorders: Opportunities and Limitations. *Genes (Basel)* **12**, doi:10.3390/genes12091427 (2021).

37 Zech, M. *et al.* Clinically relevant copy-number variants in exome sequencing data of patients with dystonia. *Parkinsonism Relat Disord* **84**, 129-134, doi:10.1016/j.parkreldis.2021.02.013 (2021).

38 Coutelier, M. *et al.* Combining callers improves the detection of copy number variants from whole-genome sequencing. *Eur J Hum Genet* **30**, 178-186, doi:10.1038/s41431-021-00983-x (2022).

39 Mok, K. Y. *et al.* Deletions at 22q11.2 in idiopathic Parkinson's disease: a combined analysis of genome-wide association data. *Lancet Neurol* **15**, 585-596, doi:10.1016/S1474-4422(16)00071-5 (2016).

40 Cunningham, A. C. *et al.* Movement Disorder Phenotypes in Children With 22q11.2 Deletion Syndrome. *Mov Disord* **35**, 1272-1274, doi:10.1002/mds.28078 (2020).

41 Pirooznia, M., Goes, F. S. & Zandi, P. P. Whole-genome CNV analysis: advances in computational approaches. *Front Genet* **6**, 138, doi:10.3389/fgene.2015.00138 (2015).

42 Lillevali, H. *et al.* Genome sequencing identifies a homozygous inversion disrupting QDPR as a cause for dihydropteridine reductase deficiency. *Mol Genet Genomic Med* **8**, e1154, doi:10.1002/mgg3.1154 (2020).

43 Chiang, T. *et al.* Atlas-CNV: a validated approach to call single-exon CNVs in the eMERGESeq gene panel. *Genet Med* **21**, 2135-2144, doi:10.1038/s41436-019-0475-4 (2019).

44 Wagner, M. *et al.* Mitochondrial DNA mutation analysis from exome sequencing-A more holistic approach in diagnostics of suspected mitochondrial disease. *J Inherit Metab Dis* **42**, 909-917, doi:10.1002/jimd.12109 (2019).

45 van der Sanden, B. *et al.* Systematic analysis of short tandem repeats in 38,095 exomes provides an additional diagnostic yield. *Genet Med* **23**, 1569-1573, doi:10.1038/s41436-021-01174-1 (2021).

46 Ibanez, K. *et al.* Whole genome sequencing for the diagnosis of neurological repeat expansion disorders in the UK: a retrospective diagnostic accuracy and prospective clinical validation study. *Lancet Neurol* **21**, 234-245, doi:10.1016/S1474-4422(21)00462-2 (2022).

47 Griffin, H. R. *et al.* Accurate mitochondrial DNA sequencing using off-target reads provides a single test to identify pathogenic point mutations. *Genet Med* **16**, 962-971, doi:10.1038/gim.2014.66 (2014).

48 Poole, O. V. *et al.* Mitochondrial DNA Analysis from Exome Sequencing Data Improves Diagnostic Yield in Neurological Diseases. *Ann Neurol* **89**, 1240-1247, doi:10.1002/ana.26063 (2021).

49 Yaldiz, B. *et al.* Twist exome capture allows for lower average sequence coverage in clinical exome sequencing. *Hum Genomics* **17**, 39, doi:10.1186/s40246-023-00485-5 (2023).

50 Dolzhenko, E. *et al.* ExpansionHunter Denovo: a computational method for locating known and novel repeat expansions in short-read sequencing data. *Genome Biol* **21**, 102, doi:10.1186/s13059-020-02017-z (2020).

51 Rafehi, H. *et al.* An intronic GAA repeat expansion in FGF14 causes the autosomal-dominant adult-onset ataxia SCA50/ATX-FGF14. *Am J Hum Genet* **110**, 105-119, doi:10.1016/j.ajhg.2022.11.015 (2023).

52 Magrinelli, F. *et al.* Detection and Characterization of a De Novo Alu Retrotransposition Event Causing NKX2-1-Related Disorder. *Mov Disord* **38**, 347-353, doi:10.1002/mds.29280 (2023).

53 Gilissen, C., Hoischen, A., Brunner, H. G. & Veltman, J. A. Disease gene identification strategies for exome sequencing. *Eur J Hum Genet* **20**, 490-497, doi:10.1038/ejhg.2011.258 (2012).

54 Boycott, K. M. *et al.* International Cooperation to Enable the Diagnosis of All Rare Genetic Diseases. *Am J Hum Genet* **100**, 695-705, doi:10.1016/j.ajhg.2017.04.003 (2017).

55 Skorvanek, M. *et al.* WARS2 mutations cause dopa-responsive early-onset parkinsonism and progressive myoclonus ataxia. *Parkinsonism Relat Disord* **94**, 54-61, doi:10.1016/j.parkreldis.2021.11.030 (2022).

56 Sleiman, S. *et al.* Compound heterozygous variants in SHQ1 are associated with a spectrum of neurological features, including early-onset dystonia. *Hum Mol Genet* **31**, 614-624, doi:10.1093/hmg/ddab247 (2022).

57 MacDonald, J. R., Ziman, R., Yuen, R. K., Feuk, L. & Scherer, S. W. The Database of Genomic Variants: a curated collection of structural variation in the human genome. *Nucleic Acids Res* **42**, D986-992, doi:10.1093/nar/gkt958 (2014).

58 Lappalainen, I. *et al.* DbVar and DGVa: public archives for genomic structural variation. *Nucleic Acids Res* **41**, D936-941, doi:10.1093/nar/gks1213 (2013).

59 Brunet, T. *et al.* De novo variants in neurodevelopmental disorders-experiences from a tertiary care center. *Clin Genet* **100**, 14-28, doi:10.1111/cge.13946 (2021).

60 Chang, F. C. *et al.* Phenotypic insights into ADCY5-associated disease. *Mov Disord* **31**, 1033-1040, doi:10.1002/mds.26598 (2016).

61 Westenberger, A. *et al.* Spectrum of FAR1 (Fatty Acyl-CoA Reductase 1) Variants and Related Neurological Conditions. *Mov Disord* **38**, 502-504, doi:10.1002/mds.29323 (2023).

62 Meyer, E. *et al.* Mutations in the histone methyltransferase gene KMT2B cause complex early-onset dystonia. *Nat Genet* **49**, 223-237, doi:10.1038/ng.3740 (2017).

63 Beetz, C. *et al.* LRRK2 Loss-of-Function Variants in Patients with Rare Diseases: No Evidence for a Phenotypic Impact. *Mov Disord* **36**, 1029-1031, doi:10.1002/mds.28452 (2021).

64 Blauwendraat, C. *et al.* Frequency of Loss of Function Variants in LRRK2 in Parkinson Disease. *JAMA Neurol* **75**, 1416-1422, doi:10.1001/jamaneurol.2018.1885 (2018).

65 Salles, P. A., Mata, I. F., Brunger, T., Lal, D. & Fernandez, H. H. ATP1A3-Related Disorders: An Ever-Expanding Clinical Spectrum. *Front Neurol* **12**, 637890, doi:10.3389/fneur.2021.637890 (2021).

66 Havrilla, J. M., Pedersen, B. S., Layer, R. M. & Quinlan, A. R. A map of constrained coding regions in the human genome. *Nat Genet* **51**, 88-95, doi:10.1038/s41588-018-0294-6 (2019).

67 Wiel, L. *et al.* MetaDome: Pathogenicity analysis of genetic variants through aggregation of homologous human protein domains. *Hum Mutat* **40**, 1030-1038, doi:10.1002/humu.23798 (2019).

68 Wiel, L. *et al.* De novo mutation hotspots in homologous protein domains identify function-altering mutations in neurodevelopmental disorders. *Am J Hum Genet* **110**, 92-104, doi:10.1016/j.ajhg.2022.12.001 (2023).

69 Singh, S. *et al.* De novo variants of NR4A2 are associated with neurodevelopmental disorder and epilepsy. *Genet Med* **22**, 1413-1417, doi:10.1038/s41436-020-0815-4 (2020).

70 Winter, B. *et al.* NR4A2 and Dystonia with Dopa Responsiveness. *Mov Disord* **36**, 2203-2204, doi:10.1002/mds.28701 (2021).

71 Jesus, S. *et al.* NR4A2 Mutations Can Cause Intellectual Disability and Language Impairment With Persistent Dystonia-Parkinsonism. *Neurol Genet* **7**, e543, doi:10.1212/NXG.0000000000000543 (2021).

72 Geisheker, M. R. *et al.* Hotspots of missense mutation identify neurodevelopmental disorder genes and functional domains. *Nat Neurosci* **20**, 1043-1051, doi:10.1038/nn.4589 (2017).

73 Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* **46**, 310-315, doi:10.1038/ng.2892 (2014).

74 Wilcox, E. H. *et al.* Evaluating the impact of in silico predictors on clinical variant classification. *Genet Med* **24**, 924-930, doi:10.1016/j.gim.2021.11.018 (2022).

75 Martin, A. R. *et al.* PanelApp crowdsources expert knowledge to establish consensus diagnostic gene panels. *Nat Genet* **51**, 1560-1565, doi:10.1038/s41588-019-0528-2 (2019).

76 Richards, S. *et al.* Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* **17**, 405-424, doi:10.1038/gim.2015.30 (2015).

77 Li, Q. & Wang, K. InterVar: Clinical Interpretation of Genetic Variants by the 2015 ACMG-AMP Guidelines. *Am J Hum Genet* **100**, 267-280, doi:10.1016/j.ajhg.2017.01.004 (2017).

78 Vears, D. F., Senecal, K. & Borry, P. Reporting practices for variants of uncertain significance from next generation sequencing technologies. *Eur J Med Genet* **60**, 553-558, doi:10.1016/j.ejmg.2017.07.016 (2017).

79 Clift, K. *et al.* Patients' views on variants of uncertain significance across indications. *J Community Genet* **11**, 139-145, doi:10.1007/s12687-019-00434-7 (2020).

80 Liu, P. *et al.* Reanalysis of Clinical Exome Sequencing Data. *N Engl J Med* **380**, 2478-2480, doi:10.1056/NEJMc1812033 (2019).

81 Mensah, N. E. *et al.* Automated reanalysis application to assist in detecting novel gene-disease associations after genome sequencing. *Genet Med* **24**, 811-820, doi:10.1016/j.gim.2021.11.021 (2022).

82 Fattahi, Z. *et al.* Iranome: A catalog of genomic variations in the Iranian population. *Hum Mutat* **40**, 1968-1984, doi:10.1002/humu.23880 (2019).

83 Cordts, I. *et al.* Adult-Onset Neurodegeneration in Nucleotide Excision Repair Disorders (NERD(ND)): Time to Move Beyond the Skin. *Mov Disord* **37**, 1707-1718, doi:10.1002/mds.29071 (2022).

84 Skorvanek, M. *et al.* Adult-Onset Neurodegeneration in Nucleotide Excision Repair Disorders: More Common than Expected. *Mov Disord* **37**, 2323-2324, doi:10.1002/mds.29245 (2022).

85 Rehm, H. L. Evolving health care through personal genomics. *Nat Rev Genet* **18**, 259-267, doi:10.1038/nrg.2016.162 (2017).

86 Gilissen, C., Hoischen, A., Brunner, H. G. & Veltman, J. A. Unlocking Mendelian disease using exome sequencing. *Genome Biol* **12**, 228, doi:10.1186/gb-2011-12-9-228 (2011).

87 Bamshad, M. J., Nickerson, D. A. & Chong, J. X. Mendelian Gene Discovery: Fast and Furious with No End in Sight. *Am J Hum Genet* **105**, 448-455, doi:10.1016/j.ajhg.2019.07.011 (2019).

88 McInnes, G. *et al.* Opportunities and challenges for the computational interpretation of rare variation in clinically important genes. *Am J Hum Genet* **108**, 535-548, doi:10.1016/j.ajhg.2021.03.003 (2021).

89 Stenson, P. D. *et al.* The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. *Hum Genet* **136**, 665-677, doi:10.1007/s00439-017-1779-6 (2017).

90 Firth, H. V. *et al.* DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. *Am J Hum Genet* **84**, 524-533, doi:10.1016/j.ajhg.2009.03.010 (2009).

91 Brandon, M. C. *et al.* MITOMAP: a human mitochondrial genome database--2004 update. *Nucleic Acids Res* **33**, D611-613, doi:10.1093/nar/gki079 (2005).

92 Lill, C. M. *et al.* Launching the movement disorders society genetic mutation database (MDSGene). *Mov Disord* **31**, 607-609, doi:10.1002/mds.26651 (2016).

93 Hindorff, L. A. *et al.* Prioritizing diversity in human genomics research. *Nat Rev Genet* **19**, 175-185, doi:10.1038/nrg.2017.89 (2018).

94 Philippakis, A. A. *et al.* The Matchmaker Exchange: a platform for rare disease gene discovery. *Hum Mutat* **36**, 915-921, doi:10.1002/humu.22858 (2015).

95 Sobreira, N., Schiettecatte, F., Valle, D. & Hamosh, A. GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. *Hum Mutat* **36**, 928-930, doi:10.1002/humu.22844 (2015).

96 Gannamani, R., van der Veen, S., van Egmond, M., de Koning, T. J. & Tijssen, M. A. J. Challenges in Clinicogenetic Correlations: One Phenotype - Many Genes. *Mov Disord Clin Pract* **8**, 311-321, doi:10.1002/mdc3.13163 (2021).

97 Neilson, D. E. *et al.* A Novel Variant of ATP5MC3 Associated with Both Dystonia and Spastic Paraplegia. *Mov Disord* **37**, 375-383, doi:10.1002/mds.28821 (2022).

98 Turner, T. N. *et al.* denovo-db: a compendium of human de novo variants. *Nucleic Acids Res* **45**, D804-D811, doi:10.1093/nar/gkw865 (2017).

99 Lappalainen, I. *et al.* The European Genome-phenome Archive of human data consented for biomedical research. *Nat Genet* **47**, 692-695, doi:10.1038/ng.3312 (2015).

100 Zurek, B. *et al.* Solve-RD: systematic pan-European data sharing and collaborative analysis to solve rare diseases. *Eur J Hum Genet* **29**, 1325-1331, doi:10.1038/s41431-021-00859-0 (2021).

101 Zech, M. *et al.* Variants in Mitochondrial ATP Synthase Cause Variable Neurologic Phenotypes. *Ann Neurol* **91**, 225-237, doi:10.1002/ana.26293 (2022).

102 Aref-Eshghi, E. *et al.* Evaluation of DNA Methylation Episignatures for Diagnosis and Phenotype Correlations in 42 Mendelian Neurodevelopmental Disorders. *Am J Hum Genet* **106**, 356-370, doi:10.1016/j.ajhg.2020.01.019 (2020).

103 Mirza-Schreiber, N. *et al.* Blood DNA methylation provides an accurate biomarker of KMT2B-related dystonia and predicts onset. *Brain*, doi:10.1093/brain/awab360 (2021).

104 Ciolfi, A. *et al.* Childhood-onset dystonia-causing KMT2B variants result in a distinctive genomic hypermethylation profile. *Clin Epigenetics* **13**, 157, doi:10.1186/s13148-021-01145-y (2021).

105 Kremer, L. S. *et al.* Genetic diagnosis of Mendelian disorders via RNA sequencing. *Nat Commun* **8**, 15824, doi:10.1038/ncomms15824 (2017).

106 Lee, H. *et al.* Diagnostic utility of transcriptome sequencing for rare Mendelian diseases. *Genet Med* **22**, 490-499, doi:10.1038/s41436-019-0672-1 (2020).

107 Yepez, V. A. *et al.* Clinical implementation of RNA sequencing for Mendelian disease diagnostics. *Genome Med* **14**, 38, doi:10.1186/s13073-022-01019-9 (2022).

108 Fresard, L. *et al.* Identification of rare-disease genes using blood transcriptome sequencing and large control cohorts. *Nat Med* **25**, 911-919, doi:10.1038/s41591-019-0457-8 (2019).

109 Amarasinghe, S. L. *et al.* Opportunities and challenges in long-read sequencing data analysis. *Genome Biol* **21**, 30, doi:10.1186/s13059-020-1935-5 (2020).

110 Miyatake, S. *et al.* Rapid and comprehensive diagnostic method for repeat expansion diseases using nanopore sequencing. *NPJ Genom Med* **7**, 62, doi:10.1038/s41525-022-00331-y (2022).

111 Wagner, N. *et al.* Aberrant splicing prediction across human tissues. *Nat Genet* **55**, 861-870, doi:10.1038/s41588-023-01373-3 (2023).

112 Magri, S. *et al.* Digenic inheritance of STUB1 variants and TBP polyglutamine expansions explains the incomplete penetrance of SCA17 and SCA48. *Genet Med* **24**, 29-40, doi:10.1016/j.gim.2021.08.003 (2022).

113 Parlar, S. C., Grenn, F. P., Kim, J. J., Baluwendraat, C. & Gan-Or, Z. Classification of GBA1 Variants in Parkinson's Disease: The GBA1-PD Browser. *Mov Disord* **38**, 489-495, doi:10.1002/mds.29314 (2023).

114 Kalogeropulou, A. F. *et al.* Impact of 100 LRRK2 variants linked to Parkinson's disease on kinase activity and microtubule binding. *Biochem J* **479**, 1759-1783, doi:10.1042/BCJ20220161 (2022).

115 Splinter, K. *et al.* Effect of Genetic Diagnosis on Patients with Previously Undiagnosed Disease. *N Engl J Med* **379**, 2131-2139, doi:10.1056/NEJMoa1714458 (2018).

116 Buphamalai, P., Kokotovic, T., Nagy, V. & Menche, J. Network analysis reveals rare disease signatures across multiple levels of biological organization. *Nat Commun* **12**, 6306, doi:10.1038/s41467-021-26674-1 (2021).

117 Bakhit, Y. *et al.* Intrafamilial and interfamilial heterogeneity of PINK1-associated Parkinson's disease in Sudan. *Parkinsonism Relat Disord* **111**, 105401, doi:10.1016/j.parkreldis.2023.105401 (2023).

118 Beijer, D. *et al.* Standards of NGS Data Sharing and Analysis in Ataxias: Recommendations by the NGS Working Group of the Ataxia Global Initiative. *Cerebellum*, doi:10.1007/s12311-023-01537-1 (2023).

119 Meneret, A. *et al.* Efficacy of Caffeine in ADCY5-Related Dyskinesia: A Retrospective Study. *Mov Disord* **37**, 1294-1298, doi:10.1002/mds.29006 (2022).

120 Gilbert, D. L., Leslie, E. J., Keddache, M. & Leslie, N. D. A novel hereditary spastic paraplegia with dystonia linked to chromosome 2q24-2q31. *Mov Disord* **24**, 364-370, doi:10.1002/mds.22363 (2009).

Acknowledgements

M.Z. and J.W. receive research support from the German Research Foundation (DFG 458949627; ZE 1213/2-1; WI 1820/14-1). MZ acknowledges grant support by the EJP

RD (EJP RD Joint Transnational Call 2022) and the German Federal Ministry of Education and Research (BMBF, Bonn, Germany), awarded to the project PreDYT (PREdictive biomarkers in DYsTonia, 01GM2302), and by the Federal Ministry of Education and Research (BMBF) and the Free State of Bavaria under the Excellence Strategy of the Federal Government and the Länder, as well as by the Technical University of Munich – Institute for Advanced Study. MZ is a member of the Medical and Scientific Advisory Council of the Dystonia Medical Research Foundation.

Author Contributions

M.Z. and J.W. designed and supervised the work. The article was written by M.Z. and J.W.

Competing Interests

The authors declare no competing interests.

Figure Legends

Figure 1 Next-generation sequencing data production and analysis workflow.
Different variant types called from individual exome or genome raw-data files are subjected to stepwise filtration, involving the integration of diverse web-based bioinformatic repositories and functional annotation sources. The filtering and prioritization steps highlighted in green are especially suitable for the analysis of single-nucleotide variants (SNVs) and short insertions and deletions (indels), but can be relevant to varying extents also during assessment of other genomic mutations (e.g., structural variants, mitochondrial DNA variants, or repeat expansions). Technical limitations need to be considered, hindering reliable detection of certain genotypic

abnormalities such as single-exon CNVs as well as larger or non-coding repeat expansions in exome sequencing data. When a narrow list of candidate genetic alterations has been identified, expert review ensues to allow for determination of rare variants for which high levels of evidence exist for association with the disease. ACMG, American College of Medical Genetics and Genomics; CNVs, copy-number variations; del, deletion; dup, duplication; gnomAD, Genome Aggregation Database; mito, mitochondrial; MME, Matchmaker Exchange; OMIM®, Online Mendelian Inheritance in Man®, pLI, probability of being loss-of-function intolerant.

Figure 2 Mutational constraint metrics aiding in variant interpretation.

(a) Sequencing information from >120,000 exomes and >15,000 genomes in the Genome Aggregation Database (gnomAD) is used to provide constraint scores for a given mutation type, such as loss-of-function variation (e.g., nonsense and splice-site mutation-inducing single-nucleotide variants [SNVs])[21]. The probability of being loss-of-function intolerant (pLI) metric (minimum score = 0, maximum score = 1.0) is calculated for each gene in gnomAD on the basis of the number of observed versus expected rare loss-of-function SNVs, taking into account the gene`s length and nucleotide-sequence context; genes with pLI scores of ≥0.9 are considered to be under severe constraint against loss-of-function mutations. An example of the loss-of-function variant-constrained gene *KMT2B* is shown, mutations of which represent an important cause of hereditary early-onset dystonia with neurodevelopmental comorbidity ("dystonia 28, childhood-onset"; OMIM:617284)[62]. In gnomAD controls, *KMT2B* has significantly fewer loss-of-function SNVs than expected (pLI score of 1.0), indicating a high degree of evolutionary selective pressure. Consistent with this, heterozygous loss-of-function *KMT2B* variants are responsible for highly-penetrant pediatric dystonia syndromes; the distribution of such mutations registered in ClinVar

is depicted below the *KMT2B* transcript scheme (ClinVar database accessed on May 10, 2023)[17]. (b) The degree of regional missense-mutation constraint can also be estimated with the help of gnomAD data[21,66,67]. For example, *NR4A2*, a gene linked to a neurodevelopmental disorder with dystonia and parkinsonism ("intellectual developmental disorder with language impairment and early-onset dopa-responsive dystonia-parkinsonism"; OMIM:619911)[69-71], contains a coding sequence with significantly fewer missense variants observed than expected; hence, missense variants mapping to this area with local missense intolerance may be regarded as high-priority candidates for disease causation. In ClinVar, disease-related missense mutations cluster for *NR4A2* in particular within this region[17], encoding a functionally important protein domain. Computational tools such as the MetaDome web server[67] offer user-friendly visualization of missense-constrained protein regions inferred from gnomAD data, as illustrated in the bottom panel; the MetaDome missense-mutation tolerance landscape of NR4A2 is shown with a schematic protein representation underneath. Specific ClinVar pathogenic and likely pathogenic *NR4A2* missense variants are included (ClinVar database accessed on May 10, 2023)[17].

Figure 3 An example of case matchmaking and disease gene discovery via the GeneMatcher platform.

A candidate ultra-rare variant in the mitochondrial complex V gene *ATP5MC3* was detected in a large dystonia/spasticity-affected pedigree from the US state Ohio[120]. Despite functional molecular characterization of the variant, the definition of a new hereditary disorder was not possible for more than a decade because no additional independent case with the same genetic defect had been identified. In 2019, the Ohio family was "matched" through the GeneMatcher[95] node to a German patient who was

found to harbor an identical *ATP5MC3* mutation. The *ATP5MC3*-related monogenic movement disorder was then firmly established[97].

Figure 4 Overview of a suggested multi-omics-based diagnostic strategy.

Multiple components of disease-causing molecular lesions can be considered in integrated multi-omics analyses, including DNA-level aberrations, disorder-specific DNA-modification patterns, as well as different layers of altered gene expression (RNA and protein). Episignatures are emerging as powerful tools to characterize the significance of variants in relation to rare phenotypes, especially those linked to defects of the epigenetic machinery. RNA-sequencing can detect different types of variant-induced transcript pathologies including aberrant expression, defective splicing, and monoallelic expression states. Quantitative proteomics is able to find protein-expression outliers caused by etiologically involved variants and to characterize associated protein complex disturbances. Although “standard“ peripheral blood-derived DNA is suitable for the study of episignatures, other patient-specific biological samples may be useful to generate accurate diagnostic output from gene-expression analyses with RNA-sequencing or proteomics experiments. Skin fibroblast cultures have been found to represent an optimal biomaterial for such multi-omics approaches[107]. Alternatively, whole-blood RNA-sequencing can be a robust strategy for the profiling of disease-relevant transcript-expression and splicing defects in patients with monogenic diseases[108]. The different analytic dimensions of multi-omics tests can be assessed separately, or, as preferred, in parallel in order to maximize improvements in molecular diagnostic yield.